This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIE.2019.2949529, IEEE Transactions on Industrial Electronics

Acoustic Source Localization From Multirotor UAVs

Daniele Salvati, Carlo Drioli, *Member, IEEE*, Giovanni Ferrin, *Member, IEEE*, and Gian Luca Foresti, *Senior Member, IEEE*

***Abstract*—We address the problem of acoustic source localization using a microphone array mounted on multirotor unmanned aerial vehicles (UAVs). Conventional localization beamforming techniques are especially challenging in these specific conditions, due to the nature and intensity of the disturbances affecting the recorded acoustic signals. The principal disturbances are related to the high frequency, narrowband noise originated by the electrical engines, and to the broadband aerodynamic noise induced by the propellers. A solution to this problem is proposed, which adopts an efficient beamforming technique for the direction of arrival (DOA) estimation of an acoustic source and a circular array detached from the multirotor vehicle body in order to reduce the effects of noise generated by the propellers. The approach used to localize the source relies on a diagonal unloading (DU) beamforming with a novel norm transform (NORT) frequency fusion. The proposed algorithm was tested on a multirotor UAV equipped with a compact uniform circular array (UCA) of eight microphones, placed on the bottom of the drone to localize the target acoustic source placed on the ground while the quadcopter is hovering at different altitudes. The experimental results conducted in outdoor hovering conditions are illustrated, and the localization performances are reported under various recording conditions and source characteristics.**

***Index Terms*—Acoustic source localization, diagonal unloading beamforming, drone, microphone array, norm transform, multirotor unmanned aerial vehicle.**

I. Introduction

Acoustic source localization (ASL), an important topic in microphone array processing since many decades, has recently proven to offer interesting application perspectives in a number of scenarios involving mobile robotic devices [1]–[5]. These include direction of arrival (DOA) estimation in a single mobile robot [6] and in mobile robot sensor networks [7], relative position estimation with an ensemble of drones [8], multimodal sound localization for humanoid robots [9], acoustic source localization for human-robot interaction [10], among others. A small number of investigations also concerned aerial acoustic scene analysis by using microphones carried by aerial drones, addressing for example relative acoustic source position estimation by a single drone [11] or drone ensembles [7]. To date, the investigation of audio array processing solutions for aerial drone applications remains however limited, despite of the wide range of applications in which environmental acoustic information would effectively complement the visual information commonly managed by unmanned aerial vehicles (UAVs). Examples of such applications are found in various domains, including civil and industrial. Civil applications include search and rescue, delivery of goods, broadcasting of sports and entertainment events, security and surveillance, agriculture, and civil infrastructure inspection [12]. Industrial application examples include energy production plant performance monitoring and power transmission line inspection [13], industrial critical structure inspections services [14], management of disasters and emergency scenarios in chemical and industrial plants [15]. Acoustic sensors carried by aerial drones are useful in a wide range of situations in which relevant information can be gathered only by acoustic sensing, but the positioning of microphones in the region of interest is impossible or impractical. Such situations are often found in the scenarios cited above. Note that acoustic sensing allows to collect acoustic-only related information through specific audio processing applications, i.e., acoustic source localization, acoustic scene analysis, source signal enhancement and remote transmission, acoustic event recognition, speech/speaker recognition. Information gathered from such acoustic data is in most cases not possible to obtain with other sensors (optical, magnetic-field, thermal, proximity), and can sometimes effectively complement their functions. E.g., when visual localization is temporarily unavailable due to occlusion or wrong camera orientation, acoustic localization information may result useful for camera steering.

The ASL problem concerns the processing of acoustic data collected by a microphone array with the aim of obtaining spatial information of the acoustic sources [16]–[22]. At today, the methods for acoustic localization can be broadly classified in two classes: indirect methods and direct methods. The indirect methods aim at estimating the time difference of the acoustic wavefront arrivals between microphone pairs [23] and then the position using geometric considerations [24]. Direct methods, on the other hand, estimate the source position of an acoustic source in a single step by exploiting some power density function representing the spatially-relevant information distribution, and they are considered in general more robust under noisy and reverberant conditions if compared to the indirect methods. The conventional steered response power (SRP) is performed with the delay and sum beamformer [25], and the minimum variance distortionless response (MVDR) [26] filter is a well-known data-dependent beamformer that provides better resolution if compared to the conventional beamformer. The multiple signal classification (MUSIC) [27]

D. Salvati, C. Drioli, G. Ferrin, and G.L. Foresti are with the Department of Mathematics, Computer Science and Physics, University of Udine, Udine 33100, Italy, e-mail: daniele.salvati@uniud.it, carlo.drioli@uniud.it, giovanni.ferrin@uniud.it, gianluca.foresti@uniud.it.

This research was partially supported by Italian MoD project a2018-045 “A proactive counter-UAV system to protect army tanks and patrols in urban areas” (Proactive_Counter_UAV).

is a high resolution and noise robust method that exploits the subspace orthogonality property to build the spatial spectrum and to localize the sources.

When the acoustic recording is performed using microphone arrays installed on multirotor aerial vehicles, the localization of acoustic sources of interest becomes especially challenging, due to the number and variety of acoustic disturbances generated by this class of devices [28]. Moreover, in the case of micro aerial vehicles (MAVs) of small size, the consequent constraints on the size of the microphone array may lead to poor sensitivity and poor spatial resolution issues. As a matter of fact, attempts to tackle the acoustic related problems typical of multirotor aerial systems have been documented only recently [11], [29]–[35]. In [11], a cross-correlation time difference of arrival (TDOA) method and a particle filter are applied to localize acoustic sources with known spectrums using an aircraft drone with one rotor. In [34], [35], methods derived from the MUSIC [27] are assessed, and the reported results show good localization performances. However, the method described also requires the monitoring of MAV inertial sensors and motor controls, and the learning or monitoring of propellers noise signal. The MUSIC method is also used in [32] with a spherical microphone array system. The conventional delay-and-sum beamformer is used in [33] with multirotor helicopters. In [29], [31], the time-frequency sparsity of specific target signals, such as speech, is exploited through the use of a time-frequency spatial filtering technique, and the method is tested in indoor laboratory prototypes. In [30], it is illustrated the performance of a localization beamforming-based spectral distance response algorithm relying on diagonal unloading (DU) beamforming, recently introduced in [36]. In the investigation, a small-size and low-cost hardware configuration is used, consisting in a 4-microphone uniform linear array of 6 cm length mounted on a micro aerial quadcopter in an indoor laboratory.

In the present study, we propose a DU beamforming with a novel norm transform (NORT) frequency fusion for the DOA estimation of an acoustic source and a new hardware arrangement, in which a uniform circular array (UCA) is placed on the bottom of the drone to localize acoustic sources at ground level while hovering, and which is detached from the multirotor vehicle body in order to reduce the effects of noise generated by the propellers.

The algorithm proposed to process the multichannel data recorded during flight is based on the narrowband DU beamforming, which provides noise robustness similar to the MUSIC method but with reduced computational cost. In fact, MUSIC method requires an eigendecomposition of the covariance matrix, while the DU beamformer is a data-dependent spatial filtering model that aims at exploiting the orthogonality property between signal and noise subspaces by subtracting an opportune diagonal matrix from the covariance matrix. The design and implementation of the DU beamformer is thus simple and effective, since it is obtained by computing the matrix (un)loading factor. A broadband localization beamformer is computed in the frequency-domain by calculating the SRP on each frequency bin and by integrating the narrowband SRP components over all frequencies. To increase the spatial resolution, the narrowband components are in general normalized with respect to some spectral characteristic. Examples are the widely used phase transform (PHAT) [23], a prefilter that uses the magnitude information of the covariance matrix to normalize the narrowband components in the SRP conventional beamforming, or the incoherent frequency fusion [37], that has been shown to increase the spatial resolution for the MUSIC, the MVDR, and the DU beamformer. In this work, we do not assume any knowledge concerning the spectral source characteristics, thus the frequency range for the computation of the DU narrowband beamforming is selected to be sufficiently wide to operate with acoustic sources that have different spectral characteristics. If the source spectrum does not span all frequencies used for the narrowband beamforming, some narrowband components are corrupted primarily by noise. To mitigate the contribution of these noisy components in the fusion, we introduce a new frequency fusion, called here NORT, which is based on the norm of the narrowband SRP. Specifically, we demonstrate that the taxicab norm (i.e., L1-norm) provides an effective broadband fusion in very high noise conditions.

With respect to other drone-specific localization techniques [11], [29]–[35], we propose a new system configuration, in which the UCA is positioned under the UAV, at a certain distance from the propellers. In this way, we significantly improve the signal-to-propeller-noise ratio (SPNR), reducing also the energy of the propellers in the acoustic map, since the UCA is mounted on a hanging circular plate and is directed towards the ground. Hence, the propeller wavefronts do not impinge directly upon the microphones. Since the SPNR affects significantly the localization performance, the detached-array configuration aims at improving the acoustic source localization by increasing the SPNR at microphones.

To summarize, the main contributions of the paper are: (1) A DU beamforming with a novel broadband NORT frequency fusion is proposed to improve the localization accuracy reducing the drone ego-noise contribution in the acoustic map and to operate with a wide range of acoustic sources with different spectral characteristics; (2) A configuration strategy in which the microphone array is detached from the drone is proposed to reduce the intensity of noise, generated by the propellers, at microphones reducing the SPNR and obtaining an effective localization in real-world conditions; (3) The DU-NORT and the detached-array configuration are validated with real-world experiments, conducted in outdoor hovering conditions at different heights, for different source target DOAs, for different sound types, and with different SPNRs.

II. Method: The DU-NORT Algorithm

A. Model

Let us refer to a UCA with M omnidirectional microphones, placed on the bottom side of the multirotor UAV, and let us address the problem of localizing an active acoustic source positioned at the ground level. We assume that the distance of the source from the array is much greater than the diameter of the UCA, consequently we will refer to a far-field model for the sound source wave propagation.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIE.2019.2949529, IEEE Transactions on Industrial Electronics

Suppose that a single source impinges upon the UCA and let $s(t)$ denote the signal generated by a nonstationary broadband source at the reference sensor and at time t. If $x_m(t)$ ($m = 1, 2, \ldots, M$) is the multichannel input captured by the array, the far-field noisy data model of the array signals in free-field can be expressed as

$$\mathbf{x}(k,f) = \mathbf{a}(f, \mathbf{\Omega}_s)S(k,f) + \mathbf{v}^d(k,f) + \mathbf{v}(k,f), \quad (1)$$

where $S(k,f)$ is the discrete-time Fourier transform (DTFT) of the source signal at the reference sensor $s(t)$, k is the block time index, f is the frequency bin, $\mathbf{v}^d(k,f)$ is the drone ego-noise, $\mathbf{v}(k,f)$ is an additive noise assumed to be spatially white Gaussian, $\mathbf{a}(f, \mathbf{\Omega}_s)$ is the array steering vector for the source direction $\mathbf{\Omega}_s = [\theta_s, \phi_s]$ (θ_s and ϕ_s are the azimuth and elevation angles), and the vectors are defined as $\mathbf{x}(k,f) = [X_1(k,f), X_2(k,f), \ldots, X_M(k,f)]^T$,$\mathbf{v}^d(k,f) = [V_1^d(k,f), V_2^d(k,f), \ldots, V_M^d(k,f)]^T$ and $\mathbf{v}(k,f) = [V_1(k,f), V_2(k,f), \ldots, V_M(k,f)]^T$, where $X_m(k,f)$, $V_m^d(k,f)$ and $V_m(k,f)$ are the DTFTs of $x_m(t)$, $v_m^d(t)$ and $v_m(t)$ respectively, and T denotes the transpose operator.

The frequency-domain model of a typical acoustic narrowband beamformer, i.e., a spatial filter whose goal is to achieve directional signal reception, can be stated as $Y(k,f,\mathbf{\Omega}) = \mathbf{w}^H(k,f,\mathbf{\Omega})\mathbf{x}(k,f)$, with $\mathbf{w}(k,f,\mathbf{\Omega})$ being the beamformer coefficients for time-shifting, weighting, and summing the data so to steer the array in the direction $\mathbf{\Omega} = [\theta, \phi]$, $Y(k,f,\mathbf{\Omega})$ being the output of the narrowband beamformer, and H denoting the conjugate transpose. The power spectral density of the spatially filtered signal is thus

$$\begin{aligned} P(k,f,\mathbf{\Omega}) &= E\{|Y(k,f,\mathbf{\Omega})|^2\} = E\{|\mathbf{w}^H(k,f,\mathbf{\Omega})\mathbf{x}(k,f)|^2\} \\ &= \mathbf{w}^H(k,f,\mathbf{\Omega})\mathbf{\Phi}(k,f)\mathbf{w}(k,f,\mathbf{\Omega}), \end{aligned} \quad (2)$$

where $|\cdot|$ denotes the absolute value, $\mathbf{\Phi}(k,f) = E\{\mathbf{x}(k,f)\mathbf{x}^H(k,f)\}$ is the covariance matrix of the array signal, and $E\{\cdot\}$ denotes mathematical expectation.

B. Narrowband DU beamforming in single-source case with spatially white noise and true covariance matrix

The DU beamformer [36] is a data-dependent spatial filtering model that aims at exploiting the orthogonality property between signal and noise subspaces by subtracting an opportune diagonal matrix from the covariance matrix $\mathbf{\Phi}(k,f)$ of the array output vector. As a result, the DU beamforming removes as much as possible the signal subspace from the covariance matrix and provides a high resolution spatial pseudo-spectrum. In practice, the design and implementation of the DU transformation is simple and effective, and is obtained by computing the matrix (un)loading factor that sets to zero the eigenvalue corresponding to the signal subspace in the theoretical model of a single source with spatially uncorrelated white Gaussian noise with zero mean and variance equal to σ^2 for all sensors. In this case, let $P_s(k,f) = E\{|S(k,f)|^2\}$ denote the power of the signal, then the covariance matrix can be written as $\mathbf{\Phi}(k,f) = P_s(k,f)\mathbf{a}(f,\mathbf{\Omega}_s)\mathbf{a}^H(f,\mathbf{\Omega}_s) + \sigma^2\mathbf{I}$, where $\mathbf{I}$ is the identity matrix.

Given the matrix $\mathbf{\Phi}(k,f)$ which represents the array output vector covariance, the DU transformed matrix can be written as

$$\mathbf{\Phi}_{\text{DU}}(k,f) = \mathbf{\Phi}(k,f) - \mu(k,f)\mathbf{I}, \quad (3)$$

where $\mu(k,f)$ is a real-valued, positive scalar, selected in such a way that the resulting matrix is negative semidefinite, that its eigenvalue corresponding to the signal subspace is null, and that its eigenvalues corresponding to the noise subspace are non-zero. The value of μ that verifies such constraints in a single source case with spatially uncorrelated white Gaussian noise can be shown to be

$$\mu(k,f) = \text{tr}[\mathbf{\Phi}(k,f)] - (M-1)\sigma^2, \quad (4)$$

where tr[$\cdot$] is the operator that computes the trace of a matrix.

The DU beamformer is formulated by using an optimization problem with an orthogonality constraint that aims to achieve the signal subspace removal and high resolution directional response. The optimization problem reads as:

$$\begin{aligned} &\text{minimize} \quad ||\mathbf{w}(k,f,\mathbf{\Omega}) - \mathbf{a}(f,\mathbf{\Omega})||_2^2, \\ &\text{subject to} \quad \mathbf{u}_s^H(k,f)\mathbf{w}(k,f,\mathbf{\Omega}) = 0, \end{aligned} \quad (5)$$

where $\mathbf{u}_s(k,f)$ is the signal subspace of $\mathbf{\Phi}(k,f)$, and $||\cdot||_2$ denotes the Euclidean norm. The DU beamformer is formulated by imposing that the spatial filter output is zero in the look direction. The minimization problem of the Euclidean square distance between the steering vector and the weight vector resides thus in the noise subspace due to the orthogonality property between signal and noise subspaces. For more details, the reader can refer to [36]. Using the method of Lagrange multipliers, the solution of (5) for the beamforming coefficients is $\mathbf{w}_{\text{DU}}(k,f,\mathbf{\Omega}) = \left(\frac{1}{\lambda}\mathbf{I}\right)\mathbf{\Phi}_{\text{DU}}(k,f)\mathbf{a}(f,\mathbf{\Omega})$, where λ is the noise eigenvalue of the matrix $\mathbf{\Phi}_{\text{DU}}(k,f)$. Substituting $\mathbf{w}_{\text{DU}}(k,f,\mathbf{\Omega})$ in (2) and considering that $\mathbf{\Phi}_{\text{DU}}(k,f) = \mathbf{\Phi}(k,f) - \mu(k,f)\mathbf{I} = \mathbf{U}\text{diag}(0,\lambda,\ldots,\lambda)\mathbf{U}^H$, where $\mathbf{U}$ is the eigenvector matrix of $\mathbf{\Phi}(k,f)$, and $\mathbf{\Phi}(k,f) = \mathbf{U}\text{diag}(MP_s(k,f) + \sigma^2, \sigma^2, \ldots, \sigma^2)\mathbf{U}^H$, we have

$$P'_{\text{DU}}(k,f,\mathbf{\Omega}) = \frac{\sigma^2}{\lambda^3}\mathbf{a}^H(f,\mathbf{\Omega})\mathbf{\Phi}_{\text{DU}}(k,f)\mathbf{a}(f,\mathbf{\Omega}), \quad (6)$$

where the quantity $\frac{\sigma^2}{\lambda^3}$ is a scalar factor that can be omitted since it has no influence on the DOA estimation.

C. Narrowband DU beamforming with drone ego-noise and estimated covariance matrix

In real-world applications, the covariance matrix $\mathbf{\Phi}(k,f)$ is unknown and it has to be estimated through the averaging of the array signal blocks [38]

$$\widehat{\mathbf{\Phi}}(k,f) = \frac{1}{B}\sum_{k_b=0}^{B-1}\mathbf{x}(k-k_b,f)\mathbf{x}^H(k-k_b,f), \quad (7)$$

where B is the number of snapshots for the averaging. There is always a certain mismatch between the estimated and the true covariance matrix, due to the finite sample size (number of snapshots), to the signal model mismatches, and to the nonstationary nature of the source. Besides that, the propeller

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIE.2019.2949529, IEEE Transactions on Industrial Electronics

noise is nonstationary and correlated at microphones, giving rise to a multisource localization problem. The solution in (4) is based on an ideal model in which a single source is corrupted by spatially white noise. This hypothesis is however easily violated in practice due to the model mismatch or when operated in multisource scenarios.

By considering a general data model with drone ego-noise (1), we can model the DU procedure taking into account an available covariance matrix. We can write the estimated eigenvalue matrix of the estimated covariance matrix $\widehat{\mathbf{\Phi}}(k,f)$ at time block k, organizing the eigenvalues of $\widehat{\mathbf{\Phi}}(k,f)$ in descending order $(\widehat{\lambda}_1 > \widehat{\lambda}_2 > \cdots > \widehat{\lambda}_M)$ as $\widehat{\mathbf{\Lambda}} = \text{diag}(\widehat{\lambda}_1, \widehat{\lambda}_2, \ldots, \widehat{\lambda}_M)$. The eigenvalue matrix of the transformed covariance matrix can be written as $\widehat{\mathbf{\Lambda}}_{\text{DU}} = \text{diag}(\widehat{\lambda}_1 - \mu(k,f), \widehat{\lambda}_2 - \mu(k,f), \ldots, \widehat{\lambda}_M - \mu(k,f))$. Assuming that the acoustic source spans the eigenvector corresponding to the largest eigenvalue $\widehat{\lambda}_1$, an effective practical DU solution is given by assuming $\mu(k,f) = \text{tr}[\widehat{\mathbf{\Phi}}(k,f)] = \widehat{\lambda}_1 + \widehat{\lambda}_2 + \cdots + \widehat{\lambda}_M$ [36]. This solution is valid for the model (1), since it guarantees that the transformed matrix $\widehat{\mathbf{\Phi}}_{\text{DU}}(k,f)$ is negative semidefinite. In fact, we have that $\text{tr}[\widehat{\mathbf{\Phi}}(k,f)] > \widehat{\lambda}_1$ ($\widehat{\lambda}_1$ is the largest eigenvalue of $\widehat{\mathbf{\Phi}}(k,f)$]), resulting in an attenuation of the signal subspace with respect to the noise subspace. Hence, the orthogonality property is exploited, even if partially, since the transformed matrix may contain a residual amount of signal subspace [36]. Since $\widehat{\mathbf{\Phi}}_{\text{DU}}(k,f) = \widehat{\mathbf{\Phi}}(k,f) - \text{tr}[\widehat{\mathbf{\Phi}}(k,f)]\mathbf{I}$ is negative semidefinite, we can write the DU pseudo-spectrum as

$$P_{\text{DU}}(k,f,\mathbf{\Omega}) = \frac{1}{\mathbf{a}^H(f,\mathbf{\Omega})[\text{tr}[\widehat{\mathbf{\Phi}}(k,f)]\mathbf{I} - \widehat{\mathbf{\Phi}}(k,f)]\mathbf{a}(f,\mathbf{\Omega})}. \tag{8}$$

D. Broadband NORT frequency fusion

Given the narrowband SRP components $P_{\text{DU}}(k,f,\mathbf{\Omega})$ (8), the corresponding broadband SRP $P(k,\mathbf{\Omega})$ is obtained by integrating the narrowband SRP over all frequencies. To increase the spatial resolution, the narrowband components are in general normalized with respect to some spectral characteristic. In [36], the incoherent frequency fusion [37] was used. The SRP $P(k,\mathbf{\Omega})$ of a beamformer conveys information on the acoustic energy coming from direction $\mathbf{\Omega}$, thus it will be characterized by a maximum peak corresponding to the source direction $\hat{\mathbf{\Omega}}_s(k)$. Therefore, the DOA estimate of the source is obtained by

$$\hat{\mathbf{\Omega}}_s(k) = \underset{\mathbf{\Omega}}{\text{argmax}}[P(k,\mathbf{\Omega})]. \tag{9}$$

The proposed norm transform (NORT) frequency fusion is defined as

$$P(k,\mathbf{\Omega}) = \sum_{f=f_{\text{min}}}^{f_{\text{max}}} \frac{P_{\text{DU}}(k,f,\mathbf{\Omega})}{||\mathbf{g}(k,f)||_p}, \tag{10}$$

where $||\cdot||_p$ denotes the p-norm (p is a real valued positive scalar) of the vector $\mathbf{g}(k,f) = [P_{\text{DU}}(k,f,\mathbf{\Omega}_1), P_{\text{DU}}(k,f,\mathbf{\Omega}_2), \ldots, P_{\text{DU}}(k,f,\mathbf{\Omega}_D)]$ that contains all the pseudo-spectrums for the considered directions D, and f_{min} and f_{max} denote the frequency range for the computation of the broadband SRP. We can note that the uniform norm proposed in [37], i.e., $p = \infty$, is the solution that corresponds to a normalization of the narrowband SRP with respect to the largest value of $\mathbf{g}(k,f)$.

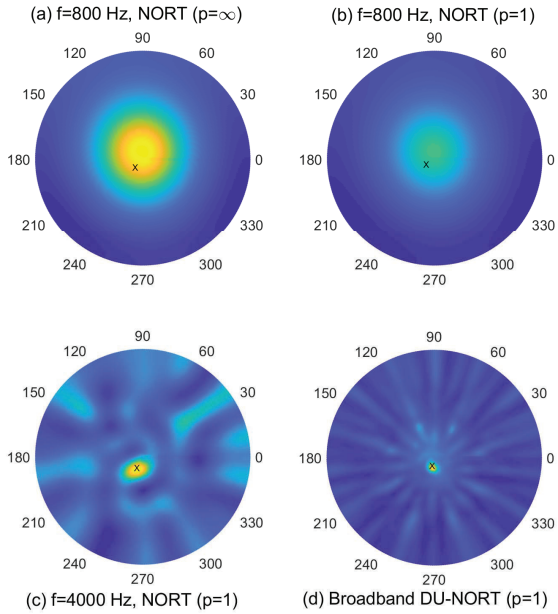

Fig. 1. Example of narrowband DU maps (a-c), and of a broadband map (d) at the same time block k computed on recorded acoustic data. The symbol x denotes the ground truth. The SPNR is about -13 dB. The source signal is a whistle sound positioned at an elevation of 5 degrees. For a frequency of 800 Hz, in which the source signal does not provide any component, we can note the amplification of the drone ego-noise due to the uniform norm ($p = \infty$) (a), and the corresponding attenuation with the taxicab norm ($p = 1$) (b). For a frequency of 4000 Hz, the source provides an active spectral component and the narrowband DU beamforming correctly estimates the DOA of the source (c). We can observe the correctly DOA estimation in the broadband fusion with the taxicab norm (d).

Under the hypothesis that the system is designed to operate with sound sources having different spectral characteristics and that the type of sources is unknown during the localization process, the broadband computation of the DU is in practical computed on a frequency range $[f_{\text{min}}, f_{\text{max}}]$ that is sufficiently wide to operate with different sound types. This requirement implies that the broadband fusion may contain narrowband components corrupted only by noise. It is clear that a normalization that assigns equal importance to each narrowband component (as the case of the NORT with $p = \infty$) introduces noise components in the broadband fusion. In very low signal-to-noise ratio (SNR) conditions, this fact can be problematic and can lead to the complete inability of estimating the source direction. Beside that, the ego-noise of a drone is composed by multiple narrowband harmonic noise originated by the electrical engines, and by the broadband aerodynamic noise induced by the propellers. The frequencies of the narrowband harmonic noise are typically nonstationary since they depend on the motor rotation speed [29]. In this scenario, the narrowband SNR, or more specifically the narrowband SPNR, varies significantly in the spectrum. A narrowband component that contains the source signal may anyhow provide a wrong information in the broadband fusion due to the low SPNR conditions.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIE.2019.2949529, IEEE Transactions on Industrial Electronics

Given these considerations, we investigate here the case $p < \infty$ instead. To better understand the improvement of the fusion using the NORT with $p < \infty$, and in particular with the L1-norm, $p = 1$, we can model the steered response power $P(f, \mathbf{\Omega})$ (k is omitted for simplicity) of a narrowband DU beamforming by considering a signal component $P_s(f, \mathbf{\Omega})$ that has a Dirac delta in the source direction with value E_s and zero value in the other directions, and a noise component $P_v(f, \mathbf{\Omega})$ due to the drone ego-noise. We have that $P(f, \mathbf{\Omega}) = P_s(f, \mathbf{\Omega}) + P_v(f, \mathbf{\Omega})$. In the noiseless case, $P_v(f, \mathbf{\Omega}) = 0, \forall \mathbf{\Omega}$, the NORT provides the same result with $p = 1$ and $p = \infty$ since $||\mathbf{g}(f)||_\infty = E_s$ and $||\mathbf{g}(f)||_1 = E_s$, and we have an SRP with value 1 in the source direction. On the other hand, when the source does not have a spectrum component in the considered frequency bin, $P_s(f, \mathbf{\Omega}_s) = 0$, we have that the max value of the power response map $\mathbf{g}(f)$ is 1 with $p = \infty$, while it is less than 1 with $p = 1$, depending on the noise distribution in the map. Since p does not affect the signal component $P_s(f, \mathbf{\Omega})$, the NORT can be formalized with the following optimization problem:

$$\begin{aligned} &\text{minimize} \quad \frac{P_v(f, \mathbf{\Omega})}{||\mathbf{g}_v(f)||_p}, \\ &\text{subject to} \quad p \geq 1. \end{aligned} \tag{11}$$

The solution is obtained with $p = 1$. An example of the NORT performance is depicted in Figure 1. The plots show three narrowband DU maps and a broadband map at the same time block k computed on acoustic data recorded by an 8-microphone UCA mounted on the bottom of a UAV. The SPNR is about -13 dB. The source signal is a whistle sound positioned at an elevation of 5 degrees. For a frequency of 800 Hz, in which the source signal does not provide any component, we can note the amplification of the drone ego-noise due to the uniform norm ($p = \infty$), and the corresponding attenuation with the taxicab norm ($p = 1$). For a frequency of 4000 Hz, the source provides an active spectral component and the narrowband DU beamforming correctly estimates the DOA of the source. We can also observe that the DOA estimation in the broadband fusion is correct.

E. Computational complexity analysis

In this section, we analyze the computational cost of the broadband DU-NORT, and we also report a comparison analysis with the SRP-PHAT [23], [25] and the broadband MUSIC [27], [37]. The computational cost is expressed in terms of the approximated number of floating-point operations (FLOPs), where a FLOP is assumed to be either a real multiplication or a real summation.

Let L denote the frame size for the fast Fourier transform (FFT), we obtain $BM(4L\log_2 L - 6L + 8)$ FLOPs for the FFTs of M channels for B snapshots. Let F denote the number of frequency bins, we obtain $M^2F(2B + 6)$ FLOPs for the estimation of covariance matrices (7). The steered response power (2) requires $FD(7M^2 + 7M - 2)$ FLOPs with D being the number of considered search directions. The sum of narrowband components has $D(F - 1)$ summations. The DU operation adds $F(M - 1)$ summations and FM subtractions to obtain the transformed matrices, and the NORT ($p = 1$) adds $F(D - 1)$ summations and FD divisions. The approximate number of FLOPs of the DU-NORT can be summarized as $BM(4L\log_2 L - 6L + 8) + M^2F(7D + 2B + 6) + MF(7D + 2) + F(D - 2) - D$. The PHAT filter in the conventional SRP requires $5FM^2$ FLOPs. We hence obtain for the SRP-PHAT a total of $BM(4L\log_2 L - 6L + 8) + M^2F(7D + 2B + 11) + 7MFD - FD - D$ FLOPs. The MUSIC instead requires an eigendecomposition that can be approximated with $13M^3$ FLOPs for the singular value decomposition of a covariance matrix [39]. The product of the noise subspace with the corresponding conjugate transpose requires $8M^3 - 8M^2 + 2M$ FLOPs. The normalized frequency fusion used in [37] adds $2FD - F$ FLOPs. The MUSIC requires approximately $BM(4L\log_2 L - 6L + 8) + 21M^3F + M^2F(7D + 2B - 2) + MF(7D + 2) + F(D - 1) - D$ FLOPs. The overall computational cost for each method is summarized in Table I. Hence, the proposed DU-NORT has a computational cost similar to the SRP-PHAT, while the MUSIC requires an eigendecomposition that has a cubic complexity of M that becomes significant at increasing of the array size. However, when the array size is small, the main contribution of the computational cost is related to the number of considered search directions. Table II shows the computational cost (FLOPs) at variation of the search directions D, considering $M = 8$, $L = 2048$, $B = 25$, $F = 635$. We can note that the DU-NORT and the SRP-PHAT provides less computational cost if compared to the MUSIC with low D, and when the number of D increases, the computational cost due to the search directions D becomes predominant, reducing the FLOPs differences between all the methods.

TABLE I
THE COMPUTATIONAL COST EXPRESSES IN TERMS OF THE APPROXIMATED NUMBER OF FLOPS.

DU-NORT
$BM(4L\log_2 L - 6L + 8) + M^2F(7D + 2B + 6) + MF(7D + 2) + F(D - 2) - D$
SRP-PHAT
$BM(4L\log_2 L - 6L + 8) + M^2F(7D + 2B + 11) + 7MFD - FD - D$
MUSIC
$BM(4L\log_2 L - 6L + 8) + 21M^3F + M^2F(7D + 2B - 2) + MF(7D + 2) + F(D - 1) - D$

TABLE II
THE COMPUTATIONAL COST (FLOPS) AT VARIATION OF THE SEARCH DIRECTIONS D FOR AN ARRAY OF 8 MICROPHONES.

D	10	100	500	1000
DU-NORT	21057870	49918530	178188130	338525130
SRP-PHAT	21239480	49985840	177747440	337449440
MUSIC	27560905	56421565	184691165	345028165

III. CONFIGURATION STRATEGY WITH THE ARRAY DETACHED FROM THE DRONE

The proposed new system configuration consists in positioning the UCA under the UAV at a certain distance from the propellers. This detached array configuration aims at improving the SPNR, and hence the localization accuracy,

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIE.2019.2949529, IEEE Transactions on Industrial Electronics

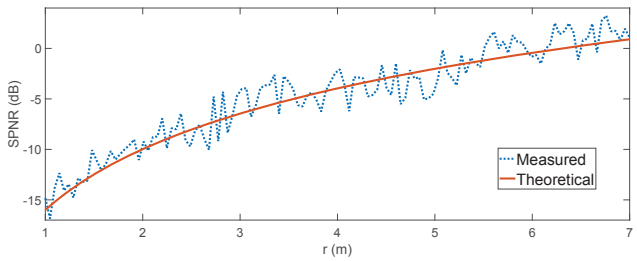

Fig. 2. Theoretical and measured SPNRs at variation of distance r between the array and the drone.

reducing also the energy of the propellers in the acoustic map, since the UCA is mounted on a hanging circular plate and is directed towards the ground. Hence, the propeller wavefronts do not impinge directly upon the microphones.

The wideband SPNR is defined as

$$\text{SPNR} = 10\log_{10}\frac{E\{\sum_{m=1}^{M}|s_m(t)|^2\}}{E\{\sum_{m=1}^{M}|v_m^d(t)|^2\}}, \tag{12}$$

where $s_m(t)$ and $v_m^d(t)$ are the time-domain m-th source signal and m-th drone ego-noise at time t and microphone m. The intensity of drone ego-noise at microphones affects significantly the localization performance, degrading the DOA estimation accuracy at very low SPNRs. By increasing the distance between the array and the drone, we can increase the SPNR providing a better localization accuracy. The effect of the distance between the array and the UAV can be theoretically analyzed with the inverse square law [40].

Said W_d the sound power of the drone ego-noise, and assuming spherical acoustical waves, the sound intensity in homogeneous and isotropic medium can be expressed as

$$I_d(r) = \frac{W_d}{4\pi r^2}, \tag{13}$$

where r is the distance from the drone. The sound intensity is thus proportional to the inverse square of the distance ($I_d(r) \propto 1/r^2$). By considering two distances r_1 and $r_2 = 2r_1$, we have that the variation ΔI of the sound intensity becomes $\Delta I = 10\log_{10}\frac{I_d(r_2)}{I_d(r_1)} = 10\log_{10}\frac{r_1^2}{(2r_1)^2} = 10\log_{10}\frac{1}{4} = -6\text{dB}$. Hence, the drone ego-noise power theoretically decreases by 6 dB each time the distance from the drone is doubled. Assuming that the distance between the acoustic source and the array is constant, the SPNR can be described by the following expression depending on the distance r

$$\text{SPNR}(r) = \text{SPNR}(r_0) - 10\log_{10}\left(\frac{r_0^2}{r^2}\right), \tag{14}$$

where $\text{SPNR}(r_0)$ is the signal-to-propeller-noise ratio for a reference distance r_0 ($r_0 < r$). Figure 2 shows the theoretical SPNR and a measured one that is computed on acoustic data recorded by an 8-microphone UCA. The source signal was a whistle sound positioned at an elevation of 0 degrees. With the hanging system removed and the array positioned at a distance of 1.7 m above the source, the drone was put vertically above the array in hovering mode, and its altitude was gradually increased so that the array-drone distance raised from 1 m to 7 m. We can see in Figure 2 that the measured SPNR follows the trend of the theoretical inverse square law.

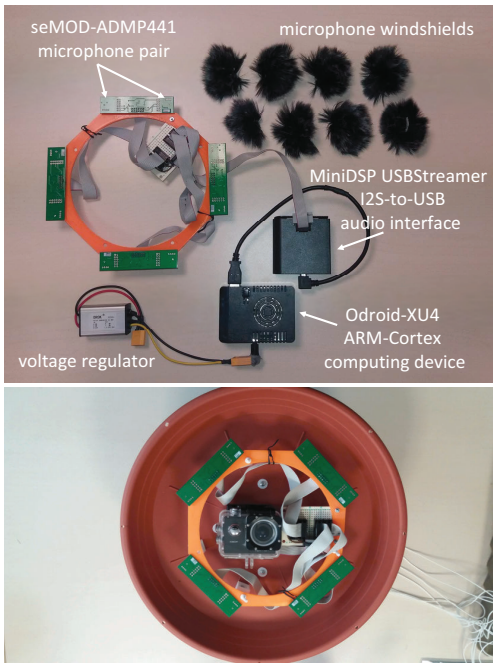

Fig. 3. The acoustic recording system used in the experiments. Top: the circular microphone array and the devices used for the signal acquisition (the 8-channel audio device, the ARM class micro-pc, the microphone windshields, and the battery pack); Bottom: the circular microphone array mounted on a hanging circular plate hosting the array itself and the audio recording devices.

IV. Experimental Setup

The multirotor UAV system selected for the study is a DJI Matrice 100 quadcopter with a 650 mm diagonal length, 2.3 kg weight, on which we mounted a compact 8-microphone UCA with a diameter of 196 mm. The microphone array was built by mounting on a circular plastic frame four Semitron seMOD-ADMP441 microphone modules, each one hosting a pair of micro electro-mechanical systems (MEMS) digital microphones in stereo configuration. The MEMS microphone has a flat frequency response from 60 Hz to 15 kHz. Each microphone pair has a distance of 40.6 mm, and the four pairs are arranged so to be equally spaced on a circumference of radius $r = 98$ mm. The MEMS capsules are covered with windshields to protect the microphone element from the wind noise. The 8 microphone channels are recorded using an Odroid-XU4 ARM-Cortex computing device, through a MiniDSP USBStreamer I2S-to-USB audio acquisition interface. The whole audio recording system is powered by a dedicated battery pack. The audio recording components are mounted on a hanging circular plate hosting the array itself, the audio recording devices (I2S interface and computing unit), and battery pack. The top of the plate is covered by polyurethane acoustic insulation foam. A picture of the recording device is provided in Figure 3

Two different configurations were investigated: setup A and setup B. In the first one, the plate with the microphones was located on the bottom of the quadcopter with a distance of

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIE.2019.2949529, IEEE Transactions on Industrial Electronics

Fig. 4. The circular microphone array mounted on the bottom of the Matrice 100 quadcopter, through a set of four nylon cords of 1 m length each (setup B).

0.25 m from the plane of the propellers, centered with respect to the four propellers. This choice is the best one in terms of compactness of the system, however it has some serious drawbacks in terms of acoustic properties, since the ego-noise of the quadcopter leads to very poor SPNRs even for small UAV-target distances. In order to mitigate the effect of the ego-noise, a different configuration was also investigated, in which the plate is hung to the quadcopter through a set of four nylon cords of 5 mm diameter and 1 m length each. [1] A picture of the UAV configured according to setup B is provided in Figure 4. In the experimental section, it will be shown how this solution leads to a sensible improvement in the SPNR of the acquired data and in the localization performance. In general, a load hanging on ropes below the drone can affect its maneuverability. This is however a situation that is encountered more and more today in a number of practical scenarios (the most important one being hauling aerial cargo), and various solutions have been made available for damping the oscillations, avoid drone swinging and improve the maneuverability in general (e.g., [41], [42]). In this study, the principal difficulties were encountered during take-off and landing, whilst during hovering and translational motion in obstacle-free space, the quadcopter was kept under control easily. A technical solution to avoid these difficulties might be to use a winch mounted below the drone to keep the sensor plate in place during take-off and landing, and to lower it once in hovering or stable flight conditions.

The audio sampling frequency was 48 kHz, and the block size was 2048 samples with a hop size of 512 samples. A Hann window was used. The covariance matrix is estimated using $B = 25$ snapshots. A spatial resolution of 2.5 degrees was used. A frequency range between 150 Hz and 15 kHz was used for broadband SRP computation, resulting 635 narrowband components. The frequency range was set considering the microphone frequency response (60 Hz to 15 kHz), and its suitability for the localization of a wide class of acoustic sources that may usually be of interest for typical acoustic scene analysis applications. These are, namely, voice sounds, ecological sounds and noises, acoustic events related to human activities and actions in a broad sense.

[1] The hanging rope system was designed as a horizontal rectangular swing hold by four parallel ropes. This ensures that when the quadcopter's attitude is horizontal, the base of the hanging system is kept horizontal. When the attitude of the UAV is not horizontal (e.g., non-null pitch to achieve constant horizontal velocity), the array plane is no more parallel to the ground, however it will still be possible to know its inclination, since it is that of the quadcopter plane. Further refinements to this design might include a gimbal system to keep the array horizontal even for non-horizontal attitudes of the UAV, however the simple hanging rope system has proved effective for the aim of this investigation.

TABLE III
THE RMSE (DEGREE) OF THE DU-NORT WITH A SCREAMING VOICE SIGNAL USING SIMULATED DATA AT VARIATION OF SPNR LEVEL. THE SNR WAS 0 dB.

SPNR (dB)	$p = 1$	$p = 2$	$p = \infty$	no norm.
-10	1.28	1.31	1.52	1.91
-11	1.44	1.46	1.95	2.61
-12	1.58	1.60	6.58	8.01
-13	2.03	2.08	11.20	12.11
-14	2.88	3.04	18.48	30.53
-15	5.24	6.50	23.20	39.41
-16	9.02	10.10	28.55	43.33
-17	14.71	18.16	35.56	44.63
-18	25.31	26.86	39.35	44.74
-19	33.56	34.66	40.27	45.43
-20	37.86	38.20	42.43	50.00

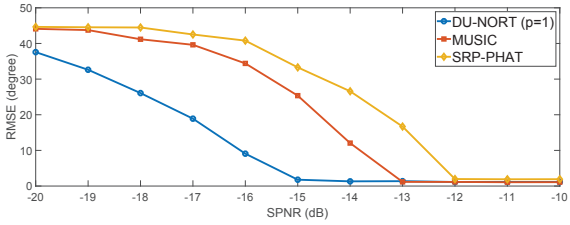

Fig. 5. The localization performance of a whistle sound signal using simulated data at variation of SPNR level. The SNR was 0 dB.

V. SIMULATIONS

In this section, we present some simulations made to test the performance of the proposed DU-NORT under real drone ego-noise conditions and spatially white Gaussian noise conditions. The simulations were conducted on a set of three sound sources: a white Gaussian noise (WGN) signal, a screaming voice, and a whistle sound. We evaluated the localization performance of the NORT using the taxicab norm ($p = 1$), the Euclidean norm ($p = 2$), and the uniform norm ($p = \infty$). We compared the DU-NORT performance with the MUSIC [27] method using the frequency fusion in [37] and with the SRP-PHAT algorithm [23], [25]. We report some simulations conducted by adding to a source signal the drone ego-noise signal recorded from a hovering UAV and by adding mutually independent white Gaussian noise to each channel. Different SPNR and SNR values where obtained by changing the ego-noise gain and the spatially white Gaussian noise level.

Table III reports the root mean square error (RMSE) of the DOA estimation of a screaming voice signal under different SPNR conditions for the DU-NORT with an SNR of 0 dB. The results show the improvement of the taxicab norm if compared to the uniform norm and to the Euclidean norm. We can also note the degradation of the performance when no normalization is used. The localization performance of the WGN source signal with an SPNR of -20 dB and an

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIE.2019.2949529, IEEE Transactions on Industrial Electronics

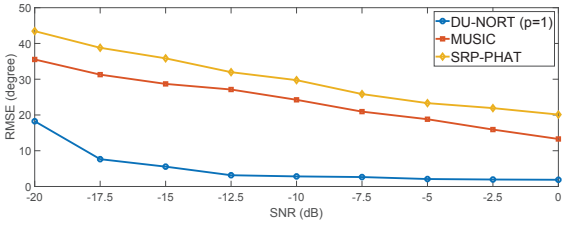

Fig. 6. The localization performance of a screaming voice signal using simulated data at variation of SNR level. The SPNR was -13 dB.

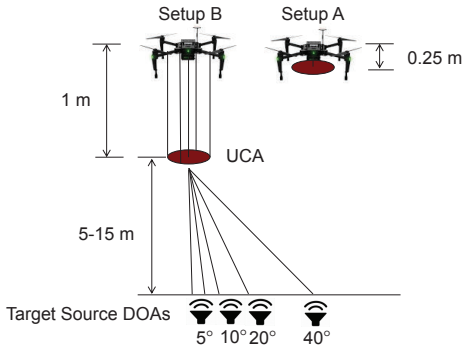

Fig. 7. Recording configuration for evaluating the setup A and B: the microphone array is on the bottom of the quadcopter for the setup A and it is 1 m below the hovering UAV for the setup B. The target acoustic source was positioned at four different angles, for three different hovering heights: 5 m, 10 m, and 15 m.

SNR of 0 dB is instead equal for all the considered norms (the RMSE is 1.6 degrees) since the frequency range for the broadband SRP is occupied by the signal in all narrowband components. Figure 5 shows the performance comparison when using a whistle sound signal. The SNR was 0 dB. The DU-NORT algorithm with the taxicab norm provides a better performance if compared to the SRP-PHAT and to the MUSIC at increasing of the noise level. Finally, Figure 6 depicts the RMSE comparison at variation of SNR levels using a screaming voice sound signal. The SPNR was -13 dB. The DU-NORT ($p = 1$) outperforms the MUSIC and the SRP-PHAT, and it is robust to the increase of the spatially white Gaussian noise level.

Hence, the taxicab norm provides a lower RMSE due to their ability in reducing the drone ego-noise in the narrowband components primarily corrupted by the noise, emphasized the target source acoustic energy in the final acoustic map. Both SRP-PHAT and MUSIC instead use a broadband fusion that assigns equal importance for each narrowband component resulting in a poor performance with the screaming voice and the whistle sound in noisy conditions.

VI. Experimental Results

The DU-NORT method described is applied to the task of localizing an acoustic source by processing the data recorded by the quadcopter equipped with the UCA discussed so far. Several recording sessions were conducted to build a database featuring different target acoustic sources at different positions with respect to the hovering UAV, and corrupted by the propeller noise in realistic acoustic conditions. In the hanging sensing plate configuration, the microphone array was positioned 1 m below the bottom of the UAV, and centered on average with respect to the four propellers. During stable hovering in the experiments, the hanging plate undergoes very small oscillations which are not influential in the DOA estimation task.[2]

The first experiment aims at comparing the setup A and setup B. The target acoustic source was generated by a loudspeaker positioned at the ground level, at different angles with respect to the UAV, namely at $5°$, $10°$, $20°$, and $40°$, with the UAV at different heights (5 m, 10 m, 15 m), as illustrated in Figure 7. The assessment was conducted on a set of three sound sources: a WGN signal of 2 seconds duration, a screaming voice of 10 seconds duration, and a whistle sound of 8 seconds duration. We use a sound pressure level meter to measure the energy of the drone and of the source and to estimate the SPNR. The measured propellers noise loudness at

TABLE IV
The RMSE (degree) of the DOA localization performance using the setup A.

WGN source			
SPNR (dB)	DU-NORT ($p = 1$)	MUSIC	SRP-PHAT
-25	35.23	38.89	40.23
-31	71.27	81.82	89.41
-34	97.01	114.14	105.54
Screaming voice source			
SPNR (dB)	DU-NORT ($p = 1$)	MUSIC	SRP-PHAT
-25	95.20	104.87	115.54
-31	98.01	106.14	118.32
-34	99.24	124.14	118.78
Whistle source			
SPNR (dB)	DU-NORT ($p = 1$)	MUSIC	SRP-PHAT
-25	96.33	104.32	110.44
-31	97.12	111.14	111.33
-34	98.01	112.44	115.72

TABLE V
The RMSE (degree) of the DOA localization performance using the setup B.

WGN source			
SPNR (dB)	DU-NORT ($p = 1$)	MUSIC	SRP-PHAT
-13	7.29	7.67	8.07
-19	10.61	10.21	11.92
-22	27.96	27.83	28.77
Screaming voice source			
SPNR (dB)	DU-NORT ($p = 1$)	MUSIC	SRP-PHAT
-13	10.56	34.60	42.22
-19	24.63	63.34	66.41
-22	91.97	107.28	97.33
Whistle source			
SPNR (dB)	DU-NORT ($p = 1$)	MUSIC	SRP-PHAT
-13	12.50	12.77	12.98
-19	32.05	74.72	70.05
-22	93.53	104.59	103.77

[2]However, the hanging plate may be subject to oscillations which, if wide, may affect the DOA estimation task. To compensate the error component due to such issue, the measurement of the relative position and orientation between the array and the drone can be addressed by using two MEMS inertial measurement units (IMUs) with integrated three-axis magnetometer, one positioned on the drone and the other on the array. The IMU has 9-degrees-of-freedom, and it achieves drift-free 3D orientation tracking with an error of 0.5 degrees [43].

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIE.2019.2949529, IEEE Transactions on Industrial Electronics

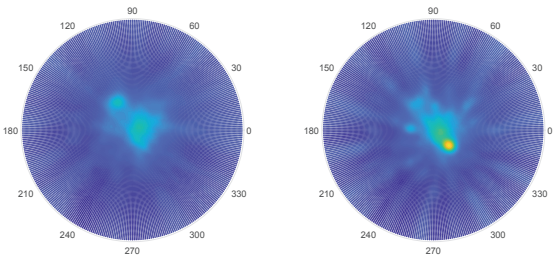

Fig. 8. The DU-NORT ($p = 1$) acoustic maps as seen at the hanging plate (setup B) in two adjacent frames of analysis. Left: the source is inactive and we can see the small energy component due to the UAV. Right: the source is active and it is clearly visible in the acoustic map. The screaming voice source was positioned at an elevation of 10 degrees. The UAV was at 10 m height. The SPNR is about -19 dB.

the array was 100 dB for the setup A and 88 dB for the setup B, and the mean loudness of the source signal at the array (with no propeller noise) was 75 dB, 69 dB and 66 dB on average, for the three different heights (5 m, 10 m, 15 m) respectively. We have that in the setup B the average sound pressure level of the UAV at the microphones is reduced by 12 dB if compared to setup A. Tables IV and V report the DOA estimate RMSE for the setup A and B, respectively. As we can observe, the RMSE is very poor for all methods, for all types of sound and for all SPNR conditions with the setup A (Table IV), except for the case at -25 dB with the WGN source. From Table V, we can see the improvement of the localization performance due to the detached array configuration. We observe that all methods have a similar performance with a WGN signal, while the DU-NORT using the taxicab norm outperforms the MUSIC and the SRP-PHAT with the voice screaming and whistle sound signal for the SPNR of -13 dB and -19 dB. When the SPNR is -22 dB the localization totally fails for all methods. Figure 8 shows the DU-NORT acoustic maps as seen at the hanging plate from two consecutive frames of analysis using the setup B. In the right plot, the source is active and it clearly visible in the acoustic map. In the left plot, the source is inactive, and we can see the small energy component due to the UAV propellers.

Next, an experiment to evaluate the localization performance for larger heights was conducted. The target acoustic source was generated by a loudspeaker positioned at the ground level with an angle of 0 degree with the UAV. The mean loudness of the source signal was 90 dB at 1 m. The UAV was positioned in stable hovering at different heights in the range [15,35] m. Table VI shows the RMSE using the setup B. The sources are correctly localized for all the heights, and we can observe that the DU-NORT provides a lower RMSE at increasing of the hovering height for the screaming voice and whistle sound signal.

Then, we have conducted an experiment to evaluate the localization performance with an interference source. The target acoustic source was generated by a loudspeaker positioned at the ground level with an angle of 0 degree with respect to the UAV. The UAV with the setup B was positioned at an hovering height of 10 m. The SPNR was -13 dB. The interference source was generated by a loudspeaker positioned at the ground level with a distance of 15 m from the target source position. The

TABLE VI
THE RMSE (DEGREE) OF THE DOA LOCALIZATION PERFORMANCE USING THE SETUP B AT VARIATION OF THE DRONE HOVERING HEIGHT. THE SOURCE IS POSITIONED WITH AN ELEVATION OF 0 DEGREES. THE MEAN LOUDNESS OF THE SOURCE WAS 90 dB AT 1 m.

WGN source			
Height (m)	DU-NORT ($p = 1$)	MUSIC	SRP-PHAT
15	0.00	0.00	0.00
20	1.44	1.44	1.44
25	1.77	1.77	1.77
30	2.50	2.50	2.50
35	2.50	2.50	2.50
Screaming voice source			
Height (m)	DU-NORT ($p = 1$)	MUSIC	SRP-PHAT
15	0.00	1.25	1.25
20	2.17	2.17	2.17
25	3.15	7.18	8.20
30	3.77	8.93	8.93
35	3.06	10.46	10.46
Whistle source			
Height (m)	DU-NORT ($p = 1$)	MUSIC	SRP-PHAT
15	1.25	1.25	6.50
20	3.95	6.37	8.75
25	4.25	7.25	10.91
30	5.23	7.23	10.40
35	6.37	7.91	10.68

TABLE VII
THE RMSE (DEGREE) OF THE DOA DU-NORT ($p = 1$) LOCALIZATION PERFORMANCE USING THE SETUP B WITH AN INTERFERENCE SOURCE. THE SPNR WAS -13 dB.

SIR (dB)	WGN	Screaming voice	Whistle
0	1.25	5.20	1.77
-10	1.25	5.20	1.77
-20	1.25	171.68	172.16

interference signal was the noise of bulldozers and digging machines at work recorded at a construction site. The mean loudness of the interference signal was set to different values to obtain three signal-to-interference ratios (SIRs): 0 dB, -10 dB, -20 dB. As we can see in Table VII, the localization of the DU-NORT fails for a SIR of -20 dB. However, the RMSE is not affected by the interference source for a SIR up -10 dB.

Last experiment was conducted with a moving UAV with the setup B. The drone was moved along a rectilinear trajectory with a hovering height of 14 m and with an average speed of 5 m/s. The UAV was first directed towards the source and then it was moved away from it. We have used a whistle sound signal. Figure 9 shows the effective localization using the DU-NORT ($p = 1$). The figure also depicts some acoustic maps in different frames and the spectrogram of a channel of the UCA. We can note the approaching to the source and the corresponding decrease of the elevation angle, and then the moving away from the source with the corresponding elevation increment.

VII. CONCLUSIONS

We have discussed the problem of acoustic source localization using a compact 8-microphone UCA installed on a quadcopter. We have presented a DU beamforming with a novel frequency fusion, called NORT, for the DOA estimation of an acoustic source. We have shown that the taxicab NORT is effective in high noise conditions when the source

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIE.2019.2949529, IEEE Transactions on Industrial Electronics

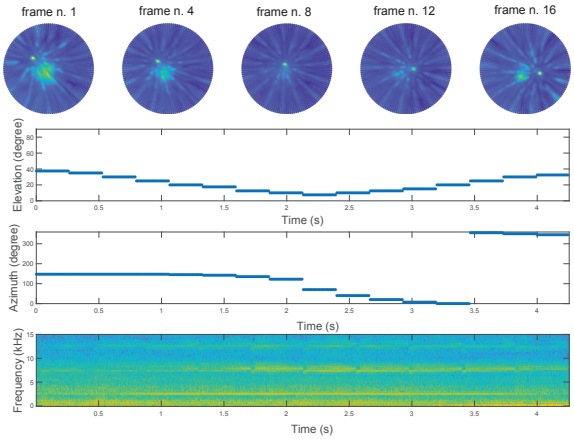

Fig. 9. The DU-NORT ($p = 1$) localization performance of a whistle signal using the setup B with the drone moving horizontally above the acoustic source.

signal spectrum does not span all the frequencies for the broadband SRP computation. We have proposed a new system configuration, in which the UCA is positioned at a certain distance under the UAV to significantly improve the SPNR at microphones and to lead an effective localization performance in realistic scenarios. Simulations and experimental results have demonstrated that the proposed system can localize successfully different types of acoustic sources up to an SPNR of about -19 dB.

REFERENCES

[1] S. Argentieri, P. Danes, and P. Soueres, "A survey on sound source localization in robotics: from binaural to array processing methods," *Computer Speech and Lang.*, vol. 34, no. 1, pp. 87–112, 2015.

[2] L. Zhang, F. Deng, J. Chen, Y. Bi, S. K. Phang, X. Chen, and B. M. Chen, "Vision-based target three-dimensional geolocation using unmanned aerial vehicles," *IEEE Trans. on Ind. Electron.*, vol. 65, no. 10, pp. 8052–8061, 2018.

[3] Y. Tang, Y. Hu, J. Cui, F. Liao, M. Lao, F. Lin, and R. S. H. Teo, "Vision-aided multi-UAV autonomous flocking in GPS-denied environment," *IEEE Trans. on Ind. Electron.*, vol. 66, no. 1, pp. 616–626, 2019.

[4] Z. Lin, H. H. T. Liu, and M. Wotton, "Kalman filter-based large-scale wildfire monitoring with a system of UAVs," *IEEE Trans. on Ind. Electron.*, vol. 66, no. 1, pp. 606–615, 2019.

[5] A. Cuenca, D. J. Antunes, A. Castillo, P. García, B. A. Khashooei, and W. P. M. H. Heemels, "Periodic event-triggered sampling and dual-rate control for a wireless networked control system with applications to UAVs," *IEEE Trans. on Ind. Electron.*, vol. 66, no. 4, pp. 3157–3166, 2019.

[6] J. M. Valin, F. Michaud, J. Rouat, and D. Letourneau, "Robust sound source localization using a microphone array on a mobile robot," in *Proc. of the IEEE/RSJ IROS*, vol. 2, pp. 1228–1233, 2003.

[7] R. Levorato and E. Pagello, "DOA acoustic source localization in mobile robot sensor networks," in *Proc.of the IEEE Int. Conf. on Auton. Robot Syst. and Comp.*, pp. 71–76, 2015.

[8] M. Basiri, F. Schill, P. Lima, and D. Floreano, "On-board relative bearing estimation for teams of drones using sound," *IEEE Robotics and Autom. Lett.*, vol. 1, no. 2, pp. 820–827, 2016.

[9] K. Youssef, S. Argentieri, and J. Zarader, "Multimodal sound localization for humanoid robots based on visio-auditive learning," in *Proc.of the IEEE Int. Conf. on Robotics and Biom.*, pp. 2517–2522, 2011.

[10] V. M. Trifa, A. Koene, J. Moren, and G. Cheng, "Real-time acoustic source localization in noisy environments for human-robot multimodal interaction," in *Proc. of th IEEE Int. Symp. on Robot and Human Inter. Comm.*, pp. 393–398, 2007.

[11] M. Basiri, F. Schill, P. U. Lima, and D. Floreano, "Robust acoustic source localization of emergency signals from micro air vehicles," in *Proc. of the IEEE/RSJ IROS*, pp. 4737–4742, 2012.

[12] S. Hayat, E. Yanmaz, and R. Muzaffar, "Survey on unmanned aerial vehicle networks for civil applications: a communications viewpoint," *IEEE Comm.s Surv. Tut.*, vol. 18, no. 4, pp. 2624–2661, 2016.

[13] F. Grimaccia, M. Aghaei, M. Mussetta, S. Leva, and P. Belezza Quate, "Planning for PV plant performance monitoring by means of unmanned aerial systems (UAS)," *Int. Journal of Energy and Env. Eng.*, vol. 6, no. 1, pp. 47–54, 2015.

[14] J. Nikolic, M. Burri, J. Rehder, S. Leutenegger, C. Huerzeler, and R. Siegwart, "A UAV system for inspection of industrial facilities," in *Proc. of the IEEE Aer. Conf.*, pp. 1–8, 2013.

[15] A. Bruzzone, F. Longo, M. Massei, L. Nicoletti, M. Agresta, R. D. Matteo, G. L. Maglione, G. Murino, and A. Padovano, "Disasters and emergency management in chemical and industrial plants: drones simulation for education and training," in *Proc. of the Int. Work. on Mod. and Sim. for Auton. Syst.*, pp. 301–308, 2016.

[16] S. Lin, "Reverberation-robust localization of speakers using distinct speech onsets and multichannel cross correlations," *IEEE/ACM Trans. on Audio, Speech, and Lang. Process.*, vol. 26, no. 11, pp. 2098–2111, 2018.

[17] H. Sundar, T. V. Sreenivas, and C. S. Seelamantula, "TDOA-based multiple acoustic source localization without association ambiguity," *IEEE/ACM Trans. on Audio, Speech, and Lang. Process.*, vol. 26, no. 11, pp. 1976–1990, 2018.

[18] D. Salvati, C. Drioli, and G. L. Foresti, "Sensitivity-based region selection in the steered response power algorithm," *Signal Process.*, vol. 153, pp. 1–100, 2018.

[19] S. Hafezi, A. H. Moore, and P. A. Naylor, "Augmented intensity vectors for direction of arrival estimation in the spherical harmonic domain," *IEEE/ACM Trans. on Audio, Speech, and Lang. Process.*, vol. 25, no. 10, pp. 1956–1968, 2017.

[20] D. Salvati, C. Drioli, and G. L. Foresti, "Exploiting a geometrically sampled grid in the steered response power algorithm for localization improvement," *J. Acoustical Soc. Amer*, vol. 141, no. 1, pp. 586–601, 2017.

[21] A. Griffin, A. Alexandridis, D. Pavlidi, Y. Mastorakis, and A. Mouchtaris, "Localizing multiple audio sources in a wireless acoustic sensor network," *Signal Process.*, vol. 107, pp. 54 –67, 2015.

[22] M. Cobos, A. Marti, and J. J. Lopez, "A modified SRP-PHAT functional for robust real-time sound source localization with scalable spatial sampling," *IEEE Signal Process. Lett.*, vol. 18, no. 1, pp. 71–74, 2011.

[23] C. Knapp and G. Carter, "The generalized correlation method for estimation of time delay," *IEEE Trans. on Acoust., Speech, and Signal Process.*, vol. 24, no. 4, pp. 320–327, 1976.

[24] P. Stoica and J. Li, "Source localization from range-difference measurements," *IEEE Signal Process. Mag.*, vol. 23, no. 3, pp. 63–66, 2006.

[25] H. Krim and M. Viberg, "Two decades of array signal processing research: the parametric approach," *IEEE Signal Process. Mag.*, vol. 13, no. 4, pp. 67–94, 1996.

[26] J. Capon, "High resolution frequency-wavenumber spectrum analysis," *Proc. of the IEEE*, vol. 57, no. 8, pp. 1408–1418, 1969.

[27] R. O. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Trans. on Antennas and Propag.*, vol. 34, no. 3, pp. 276–280, 1986.

[28] L. Wang and A. Cavallaro, "Microphone-array ego-noise reduction algorithms for auditory micro aerial vehicles," *IEEE Sensors Journal*, vol. 17, no. 8, pp. 2447–2455, 2017.

[29] L. Wang and A. Cavallaro, "Acoustic sensing from a multi-rotor drone," *IEEE Sensors Journal*, vol. 18, no. 1, pp. 4570–4582, 2018.

[30] D. Salvati, C. Drioli, G. Ferrin, and G. L. Foresti, "Beamforming-based acoustic source localization and enhancement for multirotor UAVs," in *Proc. of the EUSIPCO*, 2018.

[31] L. Wang and A. Cavallaro, "Time-frequency processing for sound source localization from a micro aerial vehicle," in *Proc. of the IEEE ICASSP*, pp. 496–500, 2017.

[32] K. Hoshiba, K. Washizaki, M. Wakabayashi, T. Ishiki, M. Kumon, Y. Bando, D. Gabriel, K. Nakadai, and H. G. Okuno, "Design of UAV-embedded microphone array system for sound source localization in outdoor environments," *Sensors*, vol. 17, no. 11, 2017.

[33] T. Ishiki and M. Kumon, "Design model of microphone arrays for multirotor helicopters," in *Proc.of the IEEE/RSJ IROS*, pp. 6143–6148, 2015.

[34] K. Furukawa, K. Okutani, K. Nagira, T. Otsuka, K. Itoyama, K. Nakadai, and H. G. Okuno, "Noise correlation matrix estimation for improving sound source localization by multirotor UAV," in *Proc.of the Int. Conf. on Intell. Robots and Syst.*, pp. 3943–3948, 2013.

[35] K. Okutani, T. Yoshida, K. Nakamura, and K. Nakadai, "Outdoor auditory scene analysis using a moving microphone array embedded in a quadrocopter," in *Proc.of the IEEE/RSJ IROS*, pp. 3288–3293, 2012.
[36] D. Salvati, C. Drioli, and G. L. Foresti, "A low-complexity robust beamforming using diagonal unloading for acoustic source localization," *IEEE/ACM Trans. on Audio, Speech, and Lang. Process.*, vol. 26, no. 3, pp. 609–622, 2018.
[37] D. Salvati, C. Drioli, and G. L. Foresti, "Incoherent frequency fusion for broadband steered response power algorithms in noisy environments," *IEEE Signal Process. Lett.*, vol. 21, no. 5, pp. 581–585, 2014.
[38] L. Zhang, W. Liu, and L. Yu, "Performance analysis for finite sample MVDR beamformer with forward backward processing," *IEEE Trans. on Signal Process.*, vol. 59, no. 5, pp. 2427–2431, 2011.
[39] L. N. Trefethen and D. Bau III, *Numerical linear algebra*. Siam, 1997.
[40] H. Kuttruff, *Room Acoustics*. Spon Press, 2009.
[41] X. Liang, Y. Fang, N. Sun, and H. Lin, "Nonlinear hierarchical control for unmanned quadrotor transportation systems," *IEEE Trans. on Ind. Electron.*, vol. 65, no. 4, pp. 3395–3405, 2018.
[42] I. Palunko, R. Fierro, and P. Cruz, "Trajectory generation for swing-free maneuvers of a quadrotor with suspended payload: A dynamic programming approach," in *Proc. of the IEEE Int. Conf. on Robotics and Autom.*, pp. 2691–2697, 2012.
[43] M. Challa and C. Wheeler, "Accuracy studies of a magnetometer-only attitude-and-rate determination system," in *Proc. of the Flight Mechanics and Est. Theory Symp.*, 1996.

Daniele Salvati is Research Fellow at the Department of Mathematics, Computer Science and Physics, University of Udine, Italy. He received the Laurea degree in environmental engineering from the Sapienza University of Rome, Italy, in 2003, the Master's degree in sound engineering from the Tor Vergata University of Rome, Italy, in 2006, and the Ph.D. degree in multimedia communication from the University of Udine, Italy, in 2012. He was a system and audio consultant to many information technology companies from 2001 to 2008. His research interests are in audio and acoustic signal processing, microphone arrays, and multimedia communications.

Carlo Drioli (M'03) received the Laurea degree in electronic engineering and the Ph.D. degree in electronic and telecommunications engineering from the University of Padova, Italy, in 1996 and 2003, respectively. He is an Assistant Professor at the University of Udine, Italy. He has been a Researcher at the Centro di Sonologia Computazionale, University of Padova; a visiting researcher at the Royal Institute of Technology (KTH), Stockholm, Sweden, with the support of the European Community through a Marie Curie Fellowship; a researcher at the Institute of Cognitive Sciences and Technology of the Italian National Research Council; a Research Assistant and Adjunct Professor at the Department of Computer Science, University of Verona. His current research interests concern multimedia signal processing, sound and voice processing by physical modeling, speech analysis and synthesis, array signal processing. He is a member of the IEEE, of the Acoustical Society of America, and of the International Speech Communication Association.

Giovanni Ferrrin received the Laurea degree cum laude in Philosophy (Natural language semantics) from Milan University and a PhD in Multimedia Communication from the University of Udine. Since 1999 he is a TA and research assistant at the Research Laboratory for New Media (NuMe) and the Artificial Vision and Real-Time Systems (AViReS) Lab (dept of Mathematics, Computer Science and Physics) University of Udine. His research interests range over the main theoretical topics in the domain of Information Fusion (context exploitation, abductive reasoning, situation awareness), Social robotics (Cognitive InfoCommunication, Ambient Assisted Living) and Educational robotics (DIY phenomena, Open Source hardware, STEM education). He is a certified UAV pilot and a member of the IEEE.

Gian Luca Foresti is Full Professor of Computer Science at the University of Udine and Director of the Dept. of Mathematics, Computer Science and Physics. He was Finance Chair of the 11th IEEE Conference on Image Processing (ICIP05), General Chair of the 16th Int. Conf. on Image Analysis and Processing (ICIAP11) and of the 8th IEEE Conf. on Advanced Video and Signal Based Surveillance (AVSS11). His main interests involve Computer Vision and Image Processing, Multisensor Data and Information Fusion, Cybersecurity, Pattern Recognition and Machine Learning. Prof. Foresti is author of more than 400 papers published in International Journals and International Conferences and he has been co-editor of several books in the field of Multimedia and Computer Vision. He has been Guest Editor of a Special Issue of the Proceedings of the IEEE on "Video Communications, Processing and Understanding for Third Generation Surveillance Systems". In 2002, he has been awarded of best IEEE Vehicular Electronics paper, in 2010 and 2019 of the Best paper Award at the International Conference on Distributed Smart Cameras (ICDSC) and in 2016 of the Best Industry Related Paper Award (BIRPA) at the International Conference on Pattern Recognition (ICPR1). Prof. Foresti is Fellow member of IAPR and Senior member of IEEE.